DOCUMENT RESUME

ED 272 356 RC 015 876

AUTHOR McConnell, Beverly B.
TITLE Basic Education for Limited English Proficient Students: The "BELEPS" Program, 1982-1985. Final Evaluation.
INSTITUTION Yakima School District Number 7, Wash.
PUB DATE 85
NOTE 57p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; *Basic Skills; *Bilingual Education Programs; Elementary Secondary Education; English (Second Language); *Hispanic Americans; Language Proficiency; *Limited English Speaking; Mathematics Achievement; Outcomes of Education; Program Descriptions; *Program Effectiveness; Program Evaluation; Reading Improvement; Second Language Instruction; Spanish Speaking; Test Results; Vocabulary Development
IDENTIFIERS *Washington (Yakima)

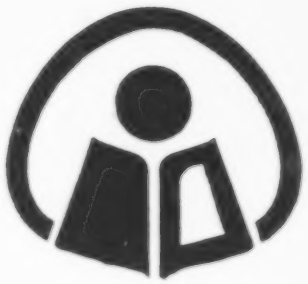
ABSTRACT

Begun in 1980 with 358 students, the Yakima (Washington) Public Schools' Basic Education for Limited English Proficient Students (BELEPS) Program served 1,250 grade K-12 students (94% Hispanic) in 1985. Its four programs were based on the degree of the students' English proficiency, and taught (1) reading only in Spanish, (2) reading in Spanish and English, (3) reading primarily in English, and (4) reading only in English. By 1985, the program employed 37 bilingual teachers. A three-year BELEPS program evaluation presented evidence of both short- and long-term gains in language proficiency and academic skills by limited English proficient students. Test instruments used included Language Assessment Skills, Peabody Picture Vocabulary Test, Wide Range Achievement Test, and Prueba de Lectura. Findings showed that Hispanic children with lowest initial proficiency in English made statistically significant gains at every grade level, indicating that children most needing the program are showing the greatest improvement. At every grade level, children showed increase in math skills. Children in dual language programs made larger and more consistent gains in reading, math, and acquisition of English vocabulary than children in programs using a modified English approach. It was recommended that findings be submitted for national validation as an educational program of "proven effectiveness." Twenty-three data tables are included. (NEC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272356

YAKIMA PUBLIC SCHOOLS
YAKIMA SCHOOL DISTRICT NUMBER 7



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Beverly B. McConnell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FINAL EVALUATION

BASIC EDUCATION FOR LIMITED ENGLISH PROFICIENT STUDENTS: THE "BELEPS" PROGRAM 1982-1985

REPORT PREPARED BY:
DR. BEVERLY B. MCCONNELL

RC015876

EVALUATION REPORT: 1982-85

FINAL REPORT ON A THREE YEAR BILINGUAL PROGRAM FOR YAKIMA PUBLIC SCHOOLS

BASIC EDUCATION FOR LIMITED ENGLISH PROFICIENT STUDENTS:

THE "BELEPS" PROGRAM

SUPERINTENDENT: DR. WARREN STARR

BILINGUAL PROJECT DIRECTOR: DR. GLORIA DE MARTINEZ

Report Prepared By:

Dr. Beverly B. McConnell

This project was partially supported by Title VII ESEA,
Washington State Bilingual Funds,
and the Yakima School District

Dr. Beverly B. McConnell
Research and Evaluation Office
SW 615 Cityview
Pullman, WA 99163

TABLE OF CONTENTS

	Page
EXECUTIVE SUMMARY	1
INTRODUCTION	5
SECTION ONE: PROJECT DESCRIPTION	8
Description of the Community and the School LEP Population	8
School Committment to Quality Education for LEP Students	9
Description of Program Models	10
SECTION TWO: THE EVALUATION DESIGN	12
SECTION THREE: STUDENT OUTCOMES	22
1. Has the BELEPS Program Improved Children's Proficiency in English?	22
2. Has the BELEPS Program Improved Children's Proficiency in Spanish?	22
3. Has the BELEPS Program Improved Children's Proficiency in English Reading?	26
4. Has the BELEPS Program Improved Children's Proficiency in Spanish Reading?	28
5. Has the BELEPS Program Improved Children's Proficiency in Math?	30
6. Is the BELEPS Program Having Different Effects With Children Who Enter With Differing Degrees or English Proficiency?	31
7. Is the BELEPS Program Having Different Effects With Children From Different Ethnic/Language Backgrounds?	34
8. Is the BELEPS Program Having Different Effects With Children Receiving Different Program Models?	39
SECTION FOUR: CONCLUSIONS AND RECOMMENDATIONS	43
APPENDICES	47

LIST OF TABLES

Table		Page
1.	English Vocabulary Scores for All BELEPS Students. . .	23
2.	Spanish Vocabulary Scores for All BELEPS Students. . .	25
3.	English Reading Scores for All BELEPS Students	27
4.	Spanish Reading by Children With Different Periods of Program Attendance	29
5.	Math Scores for All BELEPS Students.	31
6.	English Vocabulary Scores for Hispanic Students: Level A.	32
7.	English Reading Scores for Hispanic Students: Level A	33
8.	Math Scores for Hispanic Students: Level A.	34
9.	English Vocabulary Scores for All Non-Hispanic LEP Students	35
10.	English Vocabulary Scores for All Hispanic LEP Students	35
11.	English Reading Scores for All Non-Hispanic LEP Students	36
12.	English Reading Scores for All Hispanic LEP Students .	37
13.	Math Scores for All Non-Hispanic LEP Students.	38
14.	Math Scores for All Hispanic LEP Students.	39
15.	Gain in English Vocabulary by Students in Different Program Models	40
16.	Gains in English Reading by Students in Different Program Models	41
17.	Gains in Math by Students in Different Program Models .	42

APPENDICES

18.	Mean Standard Scores for All BELEPS Students by Period of Program Attendance.	47
19.	Mean Standard Scores for Hispanic BELEPS Students Who Entered the Program With The Lowest Level of English Proficiency (LAU-A).	48
20.	Mean Standard Scores for All BELEPS Students From Languages Other Than Spanish	49
21.	Mean Standard Scores for All Hispanic BELEPS Students by Period of Program Attendance.	50
22.	Mean Standard Scores of Children Enrolled in Dual Language Program Models.	51
23.	Mean Standard Scores of Children Enrolled in Modified English Instruction Program Models	52

LIST OF FIGURES

1.	Mean Standard Scores in Reading Comparing Baseline LEP Student Scores With the Scores of Children After Participation in BELEPS.	26

STUDENT PROGRESS IN YAKIMA BELEPS* PROGRAM

*BELEPS = Basic Education for Limited English Proficient Students

FINAL EVALUATION OF A THREE YEAR PROGRAM

1982 - 1985

EXECUTIVE SUMMARY

ENGLISH PROFICIENCY. Children kindergarten through grade twelve have made large and consistent gains in English vocabulary. The longer children are in the project, the higher their scores in English vocabulary. Differences based on period of attendance in the program are statistically significant in grades one through five, and at the high school level.

Hispanic children with the lowest initial proficiency in English made gains large enough to be statistically significant at every grade level, indicating that the children who need the program the most are showing the greatest improvement.

Children enrolled in dual language programs made larger and more consistent gains than children receiving modified English instruction.

SPANISH PROFICIENCY. Children at the lower elementary level, first through third grade, and in high school, made the greatest improvement in Spanish vocabulary. Gains at these levels were all statistically significant.

Spanish vocabulary scores tend to initially increase in the BELEPS program, and then to drop as children transition to more reliance on English.

ENGLISH READING. Children show improvement in English reading skills for each year of attendance in the BELEPS program. These gains are not large enough to be statistically significant in kindergarten through grade two; (reading is introduced in Spanish for many LEP children, so not all are receiving English reading instruction at these grade levels). However, from grade three through grade twelve, large and consistent gains are made in English reading skills, all of which are statistically significant. After three years in the BELEPS program, the average reading scores of children in middle school and high school are at or above national norms.

SPANISH READING. Children learning Spanish reading at first and second grade made rapid gains, with the differences between scores after one, two or three years in BELEPS statistically significant. The progress in Spanish reading is not consistent for the balance of elementary school. Children in the middle school receiving Spanish reading instruction made significant progress. At the high school level, progress was not significant. As in Spanish vocabulary, BELEPS students' scores in Spanish reading tend to increase for two years and then to drop in the third year.

MATH. At every grade level, children show increases in their math skills after attendance in the BELEPS program for one, two or

three years. Differences based on time in the program are large enough and consistent enough to be statistically significant in kindergarten through high school, except at third grade which was on the border of the limits set to establish significant gains.

After one or two years in BELEPS, children tend to score near national norms in the first through third grade. The range of scores then tends to decline in reference to national norms, even through the BELEPS students score significantly better than the baseline LEP students entering the Yakima school system at those grade levels.

DIFFERENCES BETWEEN LINGUISTIC LANGUAGE GROUPS. Approximately fifteen percent of the BELEPS students in Yakima schools have a home language other than Spanish. Separate analysis of children from the OTHER language group (non-Hispanic) found that on the average they entered the program with less English than Hispanic students, and had attained a lower level of English vocabulary after two or three years in the BELEPS program. However, gains based on program attendance were statistically significant in first grade and at all grade levels fourth grade through high school.

In English reading, children made rapid progress in elementary and middle school, with gains statistically significant at second and in fourth through eighth grades. Although the gains based on attendance were not statistically significant at third grade (probably due to the small sample size) children after two and three years in BELEPS were reading at or above national norms. "Other" language LEP children at the high school level acquired reading skills more

slowly, and, in general, attained lower scores than Hispanic LEP students.

In math, LEP children from the "other" language group made very rapid gains, statistically significant at most grade levels and, on the whole, attained consistently higher scores than Hispanic LEP students.

COMPARING PROGRAM MODELS. LEP children in dual language programs made larger and more consistent gains in reading, math and acquisition of English vocabulary than children enrolled in programs using a modified English approach. This was particularly true at the kindergarten and first grade levels. Both approaches, however, improved children's English proficiency and academic skills over the level of LEP children just entering the Yakima school system.

INTRODUCTION

This report is the third in a series evaluating the bilingual education program in the Yakima Public Schools. During the period covered by this report, Yakima Schools received federal funding under Title VII to assist with bilingual programs at certain schools heavily impacted with LEP, Limited English Proficient, students. The District also received some funding from the State of Washington to provide services to LEP students. And the District has committed its own funds in support of its policy of providing extra educational assistance to students limited in English throughout the K-12 system. This evaluation, therefore, was directed toward evaluating the progress of all LEP children, not just those in the target Title VII schools and at certain grade levels.

The purpose of the evaluation has been three-fold: 1) to provide periodic reports of student progress in specific skills areas and grade levels which have been used by the Project Director in shaping and changing the ongoing program; 2) to provide information on achievement of project goals to the funding agencies; and 3) to provide an analysis of services to LEP children and a long-range view of their educational progress over three years for the information of the Yakima community represented by the School Board members, and for school district administrators, teachers and parents of the children served on the basis of which they can make future decisions about the services to be offered to LEP children in Yakima public schools.

The information used for the compilation of this report was gathered over a three-year period of time between October, 1982 and June of 1985. Procedures for collecting data included visits to classrooms, group meetings and administration of a battery of test instruments by trained testers throughout the time period of program operation.

The format of this report will be as follows:

Section One: Project Description

This section will provide a description of the students served by the program including community variables and trends that have impacted on the program. It will also provide a description of the philosophy upon which the program has been based, and descriptions of the major program models in which children are enrolled.

Section Two: Evaluation Design

This section describes the questions which served as the focus for the evaluation of student progress, and the plan which was put in place for deriving answers to these questions. A non-traditional evaluation design has been implemented. This section explains why this design was selected in response to the unique needs of evaluating services to linguistic minority children. It also describes the procedures taken to assure validity and accuracy in the testing and analysis procedures.

Section Three: Student Outcomes

This section presents answers to the eight evaluation questions which served as the focus for the outcome evaluation. The results

of tests are reported and discussed. Several tables and figures have been developed to highlight the project findings. A more detailed statistical analysis for all of this data is included in appendices.

Section Four: Conclusions and Recommendations

This section reviews project findings and discusses policy level and programatic recommendations based on these findings.

SECTION ONE: PROJECT DESCRIPTION

Description of the Community and the School LEP Population

Yakima Metropolitan area, according to the 1980 census figures, ranked first in the state in racial minorities as a percent of the total population (17.1%). Yakima and King County together are the residence of 43% of the state's Hispanic population (26 and 25 thousand, respectively). Because of the smaller total population in Yakima County, this represents 15% of the population there, and only 2% in King County.

The minorities in Yakima County have a high birthrate and corespondingly large number of children 18 or under. Hispanics represent Washington State's "youngest" ethnic population group.

The trend is for population of ethnic and linguistic minority children to increase. In the decade between the 1970 and 1980 census, the Hispanic population in Yakima metropolitan area increased 50%. In this same period, the Asian population increased by 72%. During the five-year period from 1980 to 1985, the enrollment of LEP, Limited English Proficient students, in Yakima public schools grew as follows:

1980 - 358 students; 1981 - 440 students; 1982 - 575 students; 1983 - 893 students; 1985 - 1,250 students.

In 1985, the close of this evaluation period, Limited English Proficient students made up over 11% of the student population of 11,082 students in Yakima's public schools. Of these, 94% were

Hispanic.

Economic conditions have also been shown to affect school readiness and children's academic progress. Some relevant social and economic indicators from Census figures also reflect the social and economic conditions in Yakima. The Yakima metropolitan area ranked first in the state in the percentage of families below the poverty level, and first in the percentage of these families with children under 18. It ranked 10th in median family income.

The mobility of the student population also impacts the educational program. In 1983, school enrollment statistics showed that each of the sixteen elementary schools in Yakima had a mobility rate of over 100%, an average of 124% (total students enrolled as a percentage of year end enrollment) (YSD Evaluation Center, 1983). The elementary schools with the highest percentage of LEP students were also the schools with the highest mobility rates.

School Committment to Quality Education for LEP Students

Under the direction of Supt. Warren D. Starr, the Yakima School System has made a strong committment to providing the best possible education to limited English proficient students. They have backed this with very practical and concrete steps. One of these has been to significantly increase the number of teachers, administrators, and support personnel within the system who are bilingual. In 1980, the district had 25 bilingual teachers. In the next five years, there was a 68% increase in this figure. In 1985, at the close of this evaluation period, the district employed 37 bilingual teachers.

Bilingual education, in Yakima, has not been relegated to an "add on" part of the educational program. It has been integrated into the general basic educational program. This is reflected in the project title, BELEPS, which stands for Basic Education for Limited English Proficient Students. The District has made the existence of Spanish language editions a criteria in textbook selection.

Description of Program Models

The BELEPS students are enrolled in four basic program types. For Spanish speaking children, this is based on their degree of English proficiency and varies primarily in the degree to which the student receives native language instruction. In Program One, children would be learning reading only in Spanish; in Program Two in Spanish and English; and in Programs Three and Four, reading is taught only in English. Likewise, academic areas are taught primarily in Spanish for Program One, in either language in Program Two, primarily in English with some translation as needed in Program Three, and entirely in English in Program Four.

In all of the programs, children are being taught oral English or ESL (English as a Second Language). Most of the children who speak languages other than Spanish are enrolled in Program type Four. In this program, all instruction is in English (because the District does not have teachers who speak the variety of other languages used by this small minority of students). The classroom teacher tries to use more examples, pictures, individual attention to help a child who does not understand English well, and a pull-out teacher provides additional ESL instruction. (The District has

hired a few teachers who speak some of the "other" languages of LEP students, and has used some cross-age tutoring to provide first language communication with these students, but teaching resources are still quite limited.)

There are different staffing models used for these program types which also vary on the dimension of amount of time the student spends with someone who can understand and speak their native language. Children in Program One would mostly have both a bilingual teacher and bilingual aide, full-time. Program Two might have a bilingual teacher and part-time bilingual aide, or a monolingual (English-speaking) teacher and a bilingual aide. Program Three would have a monolingual teacher and a part-time bilingual aide or bilingual resource teacher. Program Four would have monolingual teachers and aides or resource room teachers who were also monolingual (with a few exceptions of aide and student tutors who speak some of the "other" languages, as mentioned above).

For analysis purposes, the evaluator has grouped children into two different program types. Programs One and Two are called "Dual Language Models", and Programs Three and Four are called "Modified English Instruction Models". Progress of children in each of these program models has been reported in this evaluation because this was one of the evaluation questions of interest to the District: How well do the different program models work in securing academic progress for children involved? A caution must be made--there has been no strict comparison of these two program models because they have involved children assessed as having different needs, and with

demonstrably different levels of English proficiency and initial academic performance.

The programs operate differently in the elementary schools, where most of the dual language programs are in self-contained classrooms, and at the junior and senior high school levels, where students move from class to class. At both the junior and senior high school level, the District employs some bilingual teachers who are providing instruction in Spanish to students assessed as unable to learn subject area content through strictly English presentation.

Although the children in the LEP program have been assigned to one type of program or another, there is a good deal of individualization within any program model for children with different needs. In the classrooms visited by the evaluator, there seemed to be very effective integration between the resource room personnel, the bilingual aides and classroom teachers to maximize appropriate individualized learning assistance.

SECTION TWO: EVALUATION DESIGN

The three-year evaluation of the BELEPS program was designed to answer the following evaluation questions:

1) Has the BELEPS program improved children's proficiency in English?
2) Has the BELEPS program improved children's proficiency in Spanish?
3) Has the BELEPS program improved children's proficiency in English reading?

4) Has the BELEPS program improved children's proficiency in Spanish reading?

5) Has the BELEPS program improved children's proficiency in math?

6) Is the BELEPS program having different effects with children who enter with differing degrees of English proficiency?

7) Is the BELEPS program having different effects with children from different ethnic/language backgrounds?

8) Is the BELEPS program having different effects with children receiving different program models?

Four test instruments were used to secure measurement of children's knowledge in each of these skill areas.

The LAS, Language Assessment Skills, is used to identify limited English proficient students within the District. This is one of the tests recommended for use by the Washington State Office of Bilingual Education. Spanish-speaking children are tested in both Spanish and in English; children from other home languages are tested only in English. Using the guidelines developed for this test, children are classified into five different levels of English language proficiency or "LAU" categories as defined in the Task Force Remedies LAU vs. Nichols of 1975. Category A refers to a child dominant in another language and with minimal skills in English. Category B and C are children dominant in another language with increasing knowledge of English. Category D and E are children with another home language besides English, but

with near-fluency in English (D) or full-fluency in English (E).

The Lau categories were used administratively to assign children to their initial program placement. They were used in this evaluation to classify children according to initial level of English proficiency when they entered the BELEPS program. This was then used in making comparisons between children who started with comparable levels of English language ability, after varying periods of time in the bilingual program.

The Peabody Picture Vocabulary Test, Form A in English and Form B in Spanish, was the instrument used to measure vocabulary in English and in Spanish. This test was selected because it is widely used in program evaluation nationally. It has standard scores across the entire age range covered by the bilingual program in Yakima (which enrolls children between age 5 and age 21), unlike most language tests which focus on the early school grades only. Standard scores are scores which compare the individual test performance to that of other people of the same age who took the test in the national norm group. It can also be given in much less time than most other language tests -- important because children were pulled out of classrooms for individual testing (necessary to fairly test children limited in English) and it was important to have tests that would not require excessive amounts of time. The Spanish version (Form B) was developed for use with a similar population group to that served in the Yakima schools (the Hispanic population in Yakima predominantly identifies itself as "Mexican" in ancestry).

The Wide Range Achievement Test was used to measure achievement in English reading and in math. This, too, is a nationally standardized test that has been widely used in national evaluations. The reading portion of the test is designed for individual administration, important with linguistic minority children. The test can be given in much less time than most achievement test batteries available for use in elementary and secondary schools, important for the reasons mentioned above. It has standard scores using 100 as the mean and with a standard deviation of 15 (the same scale used for standard scores on the Peabody Picture Vocabulary Test).

Spanish reading was measured using the Prueba de Lectura (Reading Test), Serie Interamericana (InterAmerican Series). This is one of the oldest and most carefully developed test series available for use with Spanish-speaking children. It has test levels that can be used to measure Spanish reading skills across the entire K-12 range included in this evaluation.

The evaluation design is based on a time-dimension evaluation model that lets program children serve as their own "controls". LEP children enter the Yakima School District every year at virtually every grade level. The scores of these children are taken to represent the probable level of achievement on that test of a LEP child without benefit of the BELEPS program. Each child is tested within 20 days of their arrival in the District--before they could have received any appreciable benefits from the BELEPS program. Since each child is given a language proficiency test at the same time as the three achievement tests, it enables the evaluator to classify the test results according to the level of initial language

proficiency.

This body of pretests, built up over the three years of project evaluation, became the baseline of scores representing how LEP children (or a given grade level, ethnic background, and initial language proficiency) would have been expected to score in the absence of the type of services offered in the BELEPS program. It becomes a local "norm group" made up of linguistic/ethnic minority children.

After entry to the BELEPS program, each child's attendance is kept individually (necessary because of the high transciency of LEP students in the Yakima District already described) and the child is re-tested after 90 days (one-half school year) and every 90 days thereafter. The half-year schedule was selected, in part, because some migrant children are in the Yakima schools less than a full school year, and this gave the program some child progress data after even a half-year of attendance. Each month, througnout the school year, as children passed a testing interval in their attendance, these children were scheduled for testing and given the battery of tests already described.

Analysis of test results uses time in program as the key variable. The presumption is that if the program is "working", children who have received more of it will perform at a higher level on tests than children who have received less. For example, a group of third grade children are tested and the tests are then sorted into groups of those who are new to BELEPS that year, others who had started at an earlier grade level and had acquired either

one, two, or three years prior attendance in BELEPS by the time they are in third grade. If the average scores of these groups of children are compared, based on period of prior attendance in BELEPS, and the scores of children with more attendance are consistently higher than those with less, you have an extremely powerful test of whether the program is working.

In most evaluations, group comparisons are made between only two time points; a pretest and a posttest. Any number of chance circumstances could affect the tests at one of these points making it appear that there was a difference or that there was not, and that this difference could be attributed to the program. When you have a series of three, or of four time measurements, however, and these consistently show, for example, that children in the program for two years do better than those with one year attendance, and at the same time do less well than those with three years attendance, the probability of this occurring by "chance" is reduced to an infinitesimal level. A series of scores, showing a positive correlation between time in the program and level of performance, is a very powerful demonstration of program effectiveness.

This evaluation design has the advantage of nullifying cultural bias that may exist in any test designed and normed with the majority culture and language, but used with a dissimilar cultural and language group. Whatever bias the test may have affects all children equally when comparisons are made between children in the same program, matched as to ethnic and language characteristics.

Most educational evaluations use a "norm referenced model" which bases it's expectation of how the children would have performed

without the program on national norms. The expectation is that if a child is at the 20th percentile in his or her standing in relationship to national norms at one age or grade level, that they will make the same amount of improvement in a given time period as children in the national norm group, and will therefore remain at the 20th percentile at the next age or grade level. This creates a crippling bias when applied to bilingual programs because research evidence has demonstrated that without special programs, minority children's scores tend to decrease in reference to national norms as they progress to higher grade levels. (This is amply illustrated in examining the relationship between national norms and the baseline score of LEP children tested at various grade levels in this evaluation--see Appendices. Entering scores tend to drop at the high grade levels.) "Norm referenced" evaluation designs, therefore, establish a false standard of what progress might have been expected in the absence of the program and to systematically underestimate (or at least warp) measures of program effectiveness.

The other common evaluation model used in education is a comparison group design. "Treatment" children are expected to show the same gains as a comparison group of "similar" children, who did not have the program, and if the "treatment" children excell the comparison group, the program is judged to be effective. In the case of bilingual programs, this design customarily fails because there is an administrative requirement (and ethical and legal reasons as well) to provide services to all children who

need it. If all children who "need" a program because of limited English proficiency are receiving it, this customarily leaves children who did not need it (fluent or native English speakers), but who are like the project children in other respects (same ethnic background, for example) to serve as a comparison group. Since the critical variable in this case is degree of proficiency in English, the comparison of groups not matched on this variable will have no validity as a measure of project outcomes. There are various statistical devices meant tc "adjust" for initial differences (analysis of co-variance) in comparing program outcomes. Since these statistical adjustments only resolve the difference in starting level, and do not take into account that the children do not have the same opportunity to benefit from instruction if one speaks English and the other does not, the results will be meaningless. Many meaningless (and misleading) evaluations have been made of bilingual programs based on this kind of statistical "magic".

Using time in program as the key variable in the evaluation design, there is no problem of lack of comparability in the comparison group--project children become their own comparison group, with children who have been in the program for one year constituting a "comparison" group to others who have been in it for two or three years, or who are just starting the program. Children who are part of the baseline group at one grade level (the grade in which they entered) become part of the "treatment" group at other grade levels. Expected "progress" is judged by the real progress of other children from the same linguistic/language minority--not the hypothetical measure of progress established by

a totally dissimilar national norm group. On the other hand, the use of standard scores (which are based on national norms) enables the findings to be related to how the progress of children relates to mainstream U.S. school children. Without using national norms as the basis of judging if a program has been effective, it is still possible to tell how minority children are performing in relationship to other U.S. school children on this particular test.

Implementation of this time-dimension evaluation design is costly. It requires collecting enough tests to provide a test comparison group for all or most grade levels. In this evaluation, analysis is based on 1,975 sets of test scores for children in grades K-12 accumulated over a three-year period. Keeping track of attendance individually to trigger testing is more time consuming than using an evaluation model that tests all children on set calendar dates. However, the results are also more meaningful in a district with high transciency, since calendar date testing can mean comparing program effects for children who have been in the classroom for two weeks along with others who have been in it all year; and not testing a large portion of the student body who may not be present at one of the testing dates (for example, migrant children who come each year at a different time than that in which tests are scheduled). Individual testing is also more costly than group testing. Carrying out this type of evaluation is therefore another example of the committment of the District, and particularly of the Project Director, Dr. Gloria de Martinez, to obtaining the most accurate measure possible of the effectiveness of the education being offered to limited English proficient children in Yakima

schools.

Children were tested by trained, bilingual testers. Teachers were not used, in order to remove any possible test bias resulting from teachers testing their own students. Testers were trained by the evaluator, who observed them in actual testing until it could be certified that they carried out the test reliably, according to the same procedures used by other testers in the program. Periodic monitoring of test effectiveness assured the continued accuracy of test results. All tests were scored by the tester, and re-scored by a second person. All test data was entered into a computer file and was cross-checked for accuracy. Analysis was done using the SPSS computer package for the social sciences at Washington State University by the evaluator. This represents, therefore, an external evaluation of program effectiveness.

The evaluator has had experience in the field of bilingual program evaluation over a period of fifteen years. Currently, she is serving on an advisory panel responsible for designing evaluation models for use in Bilingual programs nationwide. Her study, published as "Effectiveness of Bilingual Education for Migrant Students" received recognition from the National Advisory Board on Bilingual Education in 1981. The American Educational Research Association presented it's national award for best public school program evaluation in 1976 for her evaluation of the program, "Individualized Bilingual Instruction". This information is offered so that the reader will know that the evaluation design used with the Yakima program has been carefully reviewed and recognized

by professionals in education as a valid and powerful measurement of program effectiveness for use in bilingual programs.

SECTION THREE: STUDENT OUTCOMES

1. HAS THE BELEPS PROGRAM IMPROVED CHILDREN'S PROFICIENCY IN ENGLISH?

English vocabulary level was measured using the Peabody Picture Vocabulary Test, Form A, in English. As described in the section on evaluation, children's tests were accumulated over a three-year period. These tests were taken when the child entered the program and after given periods of program attendance. Analysis grouped the scores of children at each grade level according to the period of prior program attendance. As there were fewer children enrolled in the program at the upper grade levels, it became necessary to group together children in grades four and five, in grades six through eight, and grades nine through twelve. The use of standard scores makes it possible to combine scores across grade levels. (Since combining of grade levels introduces some distortion of findings, the combination of grades was kept to the minimum necessary to provide a number sufficient for statistical analysis.)

The three-year findings for all students enrolled in the BELEPS program is shown in Table 1. Examination of these results will show an extremely close and positive correlation between time in program and test score level. Note, for example, that the group of children at each grade level who had been in the BELEPS program for two years had an average score in English vocabulary that was higher than that of the group of children who had been in the program only one year.

TABLE 1

English Vocabulary Scores For
All BELEPS Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	102	52.6	93	57.0					No
1	49	65.3	123	69.4	78	75.8			Yes[a]
2	41	73.9	75	71.9	101	81.5	35	83.3	Yes[a]
3	24	75.5	70	75.5	58	77.7	35	85.1	Yes[b]
4&5	44	71.4	110	72.1	96	74.7	66	80.0	Yes[b]
6-8	66	68.5	142	72.2	95	73.5	51	76.0	No
9-12	38	62.2	71	64.9	65	72.2	22	72.7	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

It was in turn lower than the average score of children who had been in the program for three years. This is an extremely powerful finding of program effectiveness. Attendance in the BELEPS program clearly results in consistent and large increases in English vocabulary among participating students.

Analysis of variance, a statistical procedure to test whether the differences between groups on any dimension (in this case groups based on period of attendance) are large enough and reliable enough to be considered statistically significant was carried out. This is reported in the last column. In kindergarten (where it was necessary to use tests after 90 days in the program in order to have a post test comparison group at this grade level), the differences between the groups were not statistically significant. In the middle school grades, again, the differences were not statistically significant.

BEST COPY AVAILABLE

At all other grade levels, the differences were statistically significant. Significant beyond the .10 level means that the probability of this much difference between groups occurring by chance was less than 10 in 100; the .05 level means the probability was less than 5 in 100. The power of this finding is increased by the fact that these groups were not only different from one another, but that this difference was with great consistency in the same direction--longer periods of program attendance produced higher scores in English vocabulary.

2. HAS THE BELEPS PROGRAM IMPROVED CHILDREN'S PROFICIENCY IN SPANISH?

All Hispanic students were given the Peabody Picture Vocabulary Test in Spanish, using Form B of the test adapted to Spanish by the IBI program (a nationally validated bilingual program entitled, Individualized Bilingual Instruction).

The results, for children at different grade levels, grouped by period of prior program attendance are shown in Table 2. Again, at the Kindergarten level, the difference between the groups was not enough to be statistically significant. At both the first and second grade children show a very large increase in Spanish vocabulary during their first year in the program. At these grade levels, many children are receiving instruction in academic subjects in Spanish and this results in a large jump in vocabulary as academic concepts are added to the ordinary vocabulary of the home. Gains are less after that large jump, but the differences overall are great enough to be statistically significant in first and second grade.

TABLE 2

Spanish Vocabulary Scores For
All BELEPS Students

Mean Standard Scores

	N	New To BELEPS	*N*	One Year In BELEPS	*N*	Two Years In BELEPS	*N*	Three Years In BELEPS	Are Differences Significant?
K	*104*	62.0	*104*	62.9					No
1	*46*	58.7	*122*	69.1	*74*	73.6			Yes[a]
2	*35*	60.3	*70*	70.4	*92*	80.2	*31*	77.8	Yes[a]
3	*23*	68.0	*65*	74.2	*53*	73.3	*26*	77.4	No
4&5	*46*	68.3	*99*	74.3	*79*	77.2	*47*	76.3	No
6-8	*69*	81.1	*128*	81.9	*66*	88.5	*39*	83.1	No
9-12	*52*	82.1	*70*	90.9	*42*	94.8	*13*	80.6	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

Fewer children are in dual language instruction in the higher grades shown in Table 2. Children in BELEPS still show higher Spanish vocabulary scores than the baseline group on the average, at all of these grade levels.

In the upper grades a "hook" may be observed in the trend of scores--a lower score after three years in BELEPS than the scores after two years. This finding has been found in other bilingual programs using a time-dimension evaluation model. It appears to occur in areas where there is a strong community pressure to make the transition from Spanish to English, and to be related to children's increasing reliance on English for communication as they acquire proficiency in it. (This effect does not appear in high density areas of Spanish speakers, where use of Spanish for social communica-

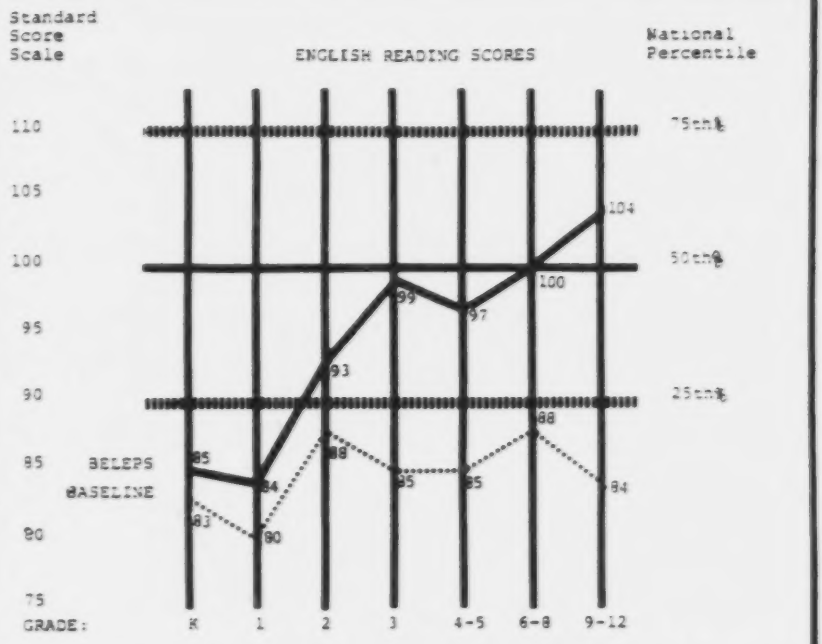
BEST COPY AVAILABLE

tion reinforces the home language even for people who have become totally fluent in English as a second language.)

Since this is a transitional bilingual program, which phases out use of Spanish as children gain proficiency in English, this result is in keeping with program policy.

3. HAS THE BELEPS PROGRAM IMPROVED CHILDREN'S PROFICIENCY IN ENGLISH READING?

The primary concern usually voiced by parents in reference to any educational program is how well children are learning to read. Figure 1 shows the difference in reading scores for children who had



NOTE: The graph shows the mean standard score for children after three years attendance for all grades except kindergarten and first grade, which represent scores after one and two years.

The solid line across represent the national norm for this test: the broken lines represent the "normal range" between the 25th and the 75th percentile of students tested nationally.

Figure 1. Mean Standard scores in reading comparing baseline LEP student scores with the scores of children after participation in BELEPS.

BEST COPY AVAILABLE

been in the BELEPS program for three years (or less at first grade and kindergarten), and the baseline scores of LEP children during the year they entered the Yakima public school system. As this Figure illustrates, the program has resulted in a very large and educationally significant gain, with children at the middle and high school level demonstrating average scores at or above the national norms for this particular test. Without the program, it would be expected that their scores in reading would have been well below the 25th percentile, as demonstrated by the average scores of the baseline group.

These test scores are shown in greater detail for all levels of attendance in Table 3. Again, it should be noted that in every case, the group of students who had been in BELEPS for two years have average scores that are higher than that of children who had been in the program only one year. These scores, in turn, are uniformly

TABLE 3

English Reading Scores For
All BELEPS Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	135	82.6	120	85.2					No
1	60	79.8	135	82.3	86	83.7			No
2	38	87.7	79	85.4	101	89.6	35	93.4	No
3	25	85.1	71	91.4	61	92.8	35	98.7	Yes[a]
4&5	51	84.8	117	90.9	98	93.2	62	96.8	Yes[a]
6-8	73	88.2	152	90.2	94	96.2	55	99.7	Yes[a]
9-12	54	84.0	95	89.6	67	99.7	25	103.7	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

BEST COPY AVAILABLE

lower than those of children who had been in BELEPS for three years. This represents a very powerful finding of effectiveness for the BELEPS program in developing children's skills in reading.

The differences at kindergarten, first and second grade are not statistically significant. Since many students begin reading in Spanish and then transition to English, these scores reflect the lack of instruction in English reading until children have been in the program longer. By third grade the differences are statistically significant, and they continue to be significant through twelfth grade. At each of these grade levels there is a relatively large gain for each attendance interval, enough that the gain in reading scores would also be considered educationally very significant.

4. HAS THE BELEPS PROGRAM IMPROVED CHILDREN'S PROFICIENCY IN SPANISH READING?

Spanish reading is taught to students who enter so dominant in Spanish that it is judged that they would be better able to begin reading in Spanish than in English. It is then phased out (as a school subject) as children acquire greater proficiency in oral English and after they have begun to progress in English reading. Their continued development of Spanish reading skills is therefore dependent on the fact that there is a large carryover in skills between reading in English and in Spanish, and the fact that the program uses text materials in other subjects which are in Spanish and in English and may be the source of additional practice in Spanish reading.

This set of circumstances fits the findings about Spanish reading skills presented in Table 4. These findings are reported in

TABLE 4

Spanish Reading* by Children With Different Periods of Program Attendance

Mean Raw Scores

Test Level	*N*	New To BELEPS	*N*	One Year In BELEPS	*N*	Two Years In BELEPS	*N*	Three Years In BELEPS	Are Differences Significant?
1	*21*	8.5	*87*	21.9	*60*	19.9	*4*	29.3	Yes[a]
2	*12*	39.9	*22*	27.2	*37*	29.6	*12*	32.8	No
3	*13*	17.4	*23*	19.9	*13*	26.8	*10*	21.5	No
4	*21*	9.2	*40*	14.3	*10*	19.9	*2*	18.5	Yes[a]
5	*17*	24.3	*29*	30.3	*7*	30.6	*1*	23.0	No

a = Significant gains beyond .05 level

NOTE: Level 1 = Grades 1,2
Level 2 = Grades 2,3
Level 3 = Grades 4,5,6
Level 4 = Grades 7,8,9
Level 5 = Grades 9,10,11

* Prueba de lectura, serie interamericana.

raw scores since there are no standard scores for Spanish reading. The Level of the test corresponds roughly to the grade levels reported below Table 4. A fairly high percentage of children took the Spanish reading test at a different level than that recommended for their grade. Because of this, the findings have been reported simply among all children who were given the test at each of the five levels shown on the Table. Most of the students in Spanish reading were at Level One. These children made large gains over the baseline group through the instruction in the BELEPS program. The correlation

BEST COPY AVAILABLE

of attendance period with higher scores is less reliable on this Table than for those presented in other subject areas. This, in part, reflects the different grade levels combined in taking a given level of the test (Level 3, for example, was taken by children ranging from grade one to grade nine).

The difference in scores based on attendance was statistically significant for Level One -- which involves most of the children in Spanish reading, and at Level 4, which represents the middle school years by and large. It was not statistically significant at the other levels. Participation in BELEPS appears to have made practically no difference in reading skills at Level 5--primarily high school students.

5. HAS THE BELEPS PROGRAM IMPROVED CHILDREN'S PROFICIENCY IN MATH?

Test results in math for children after different periods of attendance in the BELEPS program are shown in Table 5. The pattern of gains for each period of program attendance is not as consistent as in the tests of English vocabulary and English reading. However, this is the one test area in which children show a gain large enough to be statistically significant in comparison to the baseline group even at the kindergarten level. Differences based on period of attendance are also statistically significant at every grade level but one--third grade, and the probability of finding this much difference by chance was .13, very near a level that could be called statistically significant.

On a nationally normed test, children scoring between the 25th

TABLE 5

Math Scores For All
BELEPS Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	135	80.7	119	86.5					Yes[a]
1	51	90.5	136	92.7	86	97.2			Yes[a]
2	39	93.6	81	93.5	101	99.0	35	95.5	Yes[a]
3	27	89.1	71	92.2	61	91.8	35	94.7	No
4&5	52	86.2	119	84.2	100	84.7	63	88.5	Yes[b]
6-8	75	74.3	148	77.5	92	80.4	54	80.1	Yes[a]
9-12	54	77.4	95	79.4	67	86.5	25	90.8	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

and the 75th percentile are usually considered to be within the "normal" range. It should be noted that children who have attended BELEPS have scores within this low normal range in grades one through three. The scores tend to fall below this range after fourth grade. Thus, even though the BELEPS program has significantly improved children's math performance over the baseline scores, proficiency in the upper grade levels is quite low in reference to mainstream students.

6. IS THE BELEPS PROGRAM HAVING DIFFERENT EFFECTS WITH CHILDREN WHO ENTER WITH DIFFERING DEGREES OF ENGLISH PROFICIENCY?

Table 6 presents the test results for Hispanic children who entered the program at the lowest level of English proficiency, LAU

TABLE 6

English Vocabulary Scores For
Hispanic Students: Level A

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	72	41.7	72	50.1					Yes[b]
1	28	56.9	77	62.9	58	72.8			Yes[a]
2	13	66.5	41	67.5	54	79.9	18	79.1	Yes[a]
3	8	75.0	18	59.3	14	76.0	7	79.3	Yes[b]
4&5	8	44.1	20	65.1	16	72.6	10	84.6	Yes[a]
6-8	15	34.9	26	51.2	9	62.2			Yes[a]
9-12	10	33.0	19	60.3	10	62.1			Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

Level A. (There were not enough tests for children of other languages to do a separate analysis of those who entered at LAU Level A.) The gains are statistically significant at every grade level for these children who started with the lowest initial proficiency in English. This is true even with kindergarten and the middle school grades--in which the gains were not statistically significant in the analysis of scores for all BELEPS students.

In Table 7, the very low English proficiency students show a slightly different pattern of scores in English reading from the total BELEPS students. Their initial scores were so low that even the very small gains at kindergarten and first grade (in this case it would be visual recognition and matching of letters, which would be the same skill in Spanish or English) are enough to produce a "significant"

TABLE 7

English Reading Scores For
Hispanic Students: Level A

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	101	79.7	97	82.8					Yes[b]
1	27	73.7	87	80.7	63	81.9			Yes[a]
2	12	75.4	46	80.9	54	86.1	18	89.6	No
3	10	79.0	19	83.7	17	83.7	7	96.6	No
4&5	13	68.4	27	83.9	15	92.9	11	98.6	Yes[a]
6-8	21	71.7	34	80.6	10	96.6			Yes[a]
9-12	23	68.6	35	80.8	9	101.1			Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

gain. These students evidently do not get into true English reading instruction until later than the majority of BELEPS students, so their gains do not again show a significant improvement until fourth grade. From fourth grade on, the test level after two or three years was near national norms for this test. (Statistical means are not presented when the number of cases was judged to be too small to make an average score meaningful. There were not enough Level A students who took tests after three years in BELEPS at the middle school or high school level to warrant inclusion in the analysis.)

Table 8 presents the scores of LAU Level A Hispanic students in math. This mirrors the finding for the total BELEPS population (except in a lower scoring range) with statistically significant differences based on attendance at every level except third grade.

TABLE 8

Math Scores For Hispanic Students: Level A

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	101	77.3	97	84.4					Yes[a]
1	28	86.3	88	90.7	63	96.7			Yes[a]
2	12	94.2	49	91.8	54	98.8	18	94.0	Yes[a]
3	12	85.4	20	91.3	17	90.1	7	95.4	No
4&5	15	81.9	29	85.0	17	86.8	11	93.8	Yes[b]
6-8	23	71.0	33	76.6	10	80.7			Yes[a]
9-12	27	72.1	35	76.1	9	81.2			Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

In summary, the program appears to be benefiting most the children who need it the most--those students with the lowest initial level of English proficiency.

7. IS THE BELEPS PROGRAM HAVING DIFFERENT EFFECTS WITH CHILDREN FROM DIFFERENT ETHNIC/LANGUAGE BACKGROUNDS?

As noted in the description of the student population in Yakima schools, over 90% of the LEP students are Hispanic. It was not possible until the third year of the program evaluation to accumulate enough scores for separate analysis of those children who represent "other" language groups than Spanish. The majority of these children speak one of the Asian languages.

Test results in English vocabulary for children from the

"other" language group are shown in Table 9, with Table 10 presenting English vocabulary scores for all Hispanic students. The gaps in

TABLE 9

English Vocabulary Scores For All
Non-Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	6	87.0	6	75.2					No
1			9	69.8	7	81.1			Yes[b]
2			6	68.7	14	75.3			No
3			6	75.3	7	66.7	6	79.4	No
4&5			13	46.0	14	67.7	19	75.2	Yes[a]
6-8			12	50.4	24	61.0	11	69.6	Yes[a]
9-12			14	41.2	18	64.9			Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

TABLE 10

English Vocabulary Scores For
All Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	N	Are Differences Significant?
K	97	50.9	88	56.0						No
1	44	64.1	114	69.3	71	75.3				Yes[a]
2	38	76.2	69	72.2	87	82.5	32	83.6		Yes[a]
3	22	77.5	64	75.5	51	79.2	30	86.1		Yes[b]
4&5	42	73.0	97	75.6	82	75.9	48	81.8		No
6-8	63	69.1	130	74.2	71	77.7	40	77.8		Yes[a]
9-12	37	62.6	57	70.8	48	75.3	13	76.7		Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

the Table represent attendance levels where the number of tests available for analysis was too small to warrant averaging. Chance variance in a single score can produce an average that is quite misleading and for this reason, scores based on a very low number are not reported on the Tables in the text or in the Appendix. In comparing the two Tables, it can be seen that even after one year of BELEPS, the average level of English vocabulary among children of "other" languages was much lower than for the corresponding grade group of Hispanic children. Their vocabulary scores after two or three years are also lower than children in the Hispanic group at most grade levels. At the same time, the gains are very striking at the first grade, and the fourth through twelfth grade level when differences based on attendance are large enough to be statistically significant.

TABLE 11

English Reading Scores For All Non-Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	8	94.4	7	98.0					No
1	5	84.0	11	81.5	8	89.6			No
2			6	81.3	14	94.6			Yes[b]
3			6	87.2	7	101.1	5	100.2	No
4&5			13	77.1	15	88.2	17	95.8	Yes[a]
6-8			12	78.8	23	89.9	13	94.0	Yes[b]
9-12			22	83.3	19	92.0			No

41

a = Significant beyond .05 level
b = Significant beyond .10 level

In English reading, children in the "other" languages group show a significant gain based on attendance grouping at the second grade level, as shown in Table 11. Since these children begin reading in English, this is not a surprising finding. The lack of a finding of significance for reading scores at the third grade level is probably due to the lack of inclusion of baseline scores in the analysis of variance. It should be noted that the mean score of third grade children after two and after three years in the BELEPS program was above the national norm (represented by a standard score of 100). The gains made through participation in BELEPS were both large and significant at the fourth and fifth grade, and in the middle school grades. It was not enough to be significant at the high school level. Hispanic students showed significant gains from third grade through twelfth, based on attendance in the program. These results are shown in Table 12. At the upper grade levels, the average scores in reading of Hispanic students tended to be higher than the children from "other" language groups.

TABLE 12

English Reading Scores For
All Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	127	81.8	113	84.4					No
1	44	78.8	124	82.4	78	83.2			No
2	36	86.9	73	85.8	87	88.8	32	92.2	No
3	24	84.8	65	91.8	54	91.7	30	98.4	Yes[b]
4&5	47	87.2	104	92.6	83	94.2	45	97.2	Yes[a]
6-8	70	88.5	140	91.2	71	98.3	42	101.4	Yes[a]
9-12	49	84.0	74	91.5	47	103.3	14	107.4	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

Math scores for the "other" language groups are shown in Table 13, with the corresponding results for Spanish speaking children in Table 14. Unlike English vocabulary and reading, children from the "other" language groups tend to score consistently higher in math than the Hispanic children. The finding of fewer grade levels where the differences based on attendance were statistically significant is most probably due to the small number of cases (it is much more difficult to reach a level of statistical significance with very

TABLE 13

Math Scores For All
Non-Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	*N*	One Year In BELEPS	*N*	Two Years In BELEPS	*N*	Three Years In BELEPS	Are Differences Significant?
K	*8*	95.6	*7*	97.4					No
1	*5*	87.0	*11*	90.2	*8*	100.9			No
2			*6*	91.0	*14*	100.9			Yes[b]
3			*6*	89.3	*7*	94.6	*5*	98.0	No
4&5			*13*	77.6	*15*	85.7	*18*	90.7	Yes[a]
6-8			*12*	77.0	*23*	86.9	*13*	86.7	Yes[b]
9-12			*21*	85.3	*19*	94.9			Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

TABLE 14

Math Scores For All
Hispanic LEP Students

Mean Standard Scores

	N	New To BELEPS	N	One Year In BELEPS	N	Two Years In BELEPS	N	Three Years In BELEPS	Are Differences Significant?
K	127	79.7	112	85.9					Yes[a]
1	45	90.3	125	92.9	78	96.8			Yes[a]
2	36	94.7	75	93.7	87	98.8	32	94.9	Yes[a]
3	26	88.7	65	92.5	54	91.4	30	94.1	No
4&5	48	87.0	106	85.0	85	84.6	45	87.7	No
6-8	72	73.5	136	77.6	69	78.3	41	78.0	Yes[a]
9-12	53	76.9	74	77.8	47	83.0	13	85.5	Yes[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

small numbers of cases). For the attendance levels in which there were enough cases for calculation of means, there is a perfect correlation between time in the program and increases in math performance. For both groups, the BELEPS program produced large and significant learning gains.

8. IS THE BELEPS PROGRAM HAVING DIFFERENT EFFECTS WITH CHILDREN RECEIVING DIFFERENT PROGRAM MODELS?

For purpose of analysis of test scores of children enrolled in the different types of programs (described in Section One of this report), the two teaching models which involve Spanish reading, and instruction in Spanish in some other curricular areas are grouped together as the "Dual Language Models". Program Three, which

teaches reading only in English and only uses occasional translation into the primary language, is grouped with Program Four, relying entirely on English as the "Modified English Instruction" models.

The three Tables in this section show the comparison between these two models interms of whether program attendance produced no gains (e.g., scores were not, on the average, higher than the baseline group for children after different periods of program attendance); moderate gains (e.g., an increase in scores based on attendance but not large enough to be statistically significant); or, significant gains (e.g., increases large enough and consistent enough with longer periods of program attendance to be statistically significant).

TABLE 15

Gain in English Vocabulary by Students in Different Program Models

	Dual Language Instruction			Modified English Instruction		
	Significant Gains?			Significant Gains?		
	None	Moderate	Significant	None	Moderate	Significant
K		X			X	
1			X[a]	X		
2			X[a]			X[b]
3		X			X	
4&5			X[b]		X	
6-8			X[b]		X	
9-12			X[a]	X		

a = Significant beyond .05 level
b = Significant beyond .10 level

Table 15 compares scores in English vocabulary for children in the two program models. Children in dual language instruction made larger gains, which were more likely to be statistically significant than those in the modified English instruction. As shown in Table 16, there was relatively little difference between the models in English reading, insofar as both produced consistent gains. Children in dual language instruction did not show statistically significant gains until after sixth grade.

TABLE 16

Gains in English Reading by Students in Different Program Models

	Dual Language Instruction			Modified English Instruction		
	Significant Gains?			Significant Gains?		
	None	Moderate	Significant	None	Moderate	Significant
K		X			X	
1		X			X	
2		X			X	
3		X				X[b]
4&5		X			X	
6-8			X[b]		X	
9-12			X[a]			X[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

Table 17 compares the two program models in math. Again, the dual language instruction appears to produce more consistent and more significant gains than the children receiving modified English instruction.

The children assigned to dual language instruction were those who entered with the lowest level of English proficiency and

TABLE 17

Gains in Math by Students in
Different Program Models

	Dual Language Instruction			Modified English Instruction		
	Significant Gains?			Significant Gains?		
	None	Moderate	Significant	None	Moderate	Significant
K			X[a]		X	
1			X[a]		X	
2			X[a]		X	
3			X[a]	X		
4&5		X		X		
6-8			X[a]			X[a]
9-12		X				X[a]

a = Significant beyond .05 level
b = Significant beyond .10 level

corresponding low levels of performance in academic areas. Children assigned to the modified English instruction model were those with higher initial levels of English proficiency, and those who had been "transitioned" out of a dual language program because it was felt they were ready to progress with primarily English instruction. This was not always the case with the children from the "other" language group who might enter with a very low level of English proficiency but be assigned to this program model for lack of another alternative.

Because the two program models served children with different needs, reflecting a different range of test performance, these results should be taken with some caution. Undoubtedly, this is evidence that the dual language instruction model is working very well for the children assigned to it. Both models, however, have

contributed significantly to the educational progress of LEP children.

SECTION FOUR: CONCLUSIONS AND RECOMMENDATIONS

This evaluation has presented evidence of both short and longer term gains in language proficiency and academic skills by LEP students participating in Yakima's BELEPS program. This test evidence shows large and continuing growth in English proficiency, both in vocabulary and reading skills for participating children. It also has shown that children are making significant improvement in math skills at nearly every grade level.

Children at the beginning level of Spanish reading are making rapid progress in first and second garde. This is less consistently the case at higher grade levels, possibly due to the lack of a consistent policy as regards the purpose of Spanish reading instruction. There is little motivation to acquire a level of proficiency in Spanish reading when it is dropped as an academic subject after the child has been started with reading in English.

The BELEPS program seems to be effective for children from "other" language groups as well as for Hispanic children. Children from the "other" language groups tend to score lower than Hispanic LEP students in English vocabulary at all grade levels, and in English reading at the upper grade levels; to score higher than Hispanics in math at all grade levels.

The children who need the program the most, those with the lowest level of initial proficiency in English, are showing the

greatest gains.

Children in dual language instruction program models tend to make more consistent gains and more gains that are statistically significant than children in a modified English program model. However, both programs are effective overall, and have contributed greatly to the academic progress of LEP children. The two program models cannot be strictly compared because they serve children with different needs, who enter the program with different levels of English and academic proficiency.

RECOMMENDATIONS

The evaluator recommends that the Yakima School District consider submitting the findings of program effectiveness to the Joint Dissemination Review Panel of the U.S. Office of Education for possible national validation as an educational program of "proven effectiveness". It seems to possess all of the requirements for such recognition. To be validated the program must have some means of establishing how project children might have performed without benefit of the educational intervention. The accumulated scores of children entering the program are a reasonable basis for judging that this represents the probable level of performance of LEP children based on their experience in other school systems and their home environment and individual aptitudes--without some form of special intervention. The findings of increased test performance with each successive period of attendance in the BELEPS program in test after test is convincing proof of educational effectiveness, unquestionably linked to program attendance.

In addition to the test data and analysis showing statistical significance of student test scores, the School District should augment their presentation with other evidence of improved academic prospects for LEP children during the period of project implementation covered by this evaluation. One such indicator would be the increasing numbers of Hispanic students at higher grade levels. In 1979-80, there were no Hispanic students graduating from high school. In 1984-85, there were 15. In 1979-80, there were 23 Hispanic students finishing 9th grade and enrolling in the high school. In 1984-85, there were 75. In 1979-80, there were 14 Hispanic students finishing the 6th grade and enrolling in Junior High School. In 1984-85, there were 59. These are all valid indicators of the effectiveness of educational improvements implemented for Hispanic students through the BELEPS program.

Whether or not the District would wish to take on the burden of disseminating effective practices to other school districts, it would be significant to have this program receive the recognition of validation. There are very few bilingual programs which have received national recognition and validation. This program has undertaken a very extensive evaluation effort; it has the type of evidence of effectiveness required by the validation review process. The policy in support of bilingual education by the Board, the administrators, and teaching staff have produced a quality program.

APPENDICES

Table 18

Mean Standard Scores For All BELEPS Students
by Period of Program Attendance

Grade	Attend. Group	N	English Vocab.	N	Spanish Vocab.	N	English Reading	N	Math
K	0	102	52.6	104	62.0	135	82.6	135	80.7
	1	93	57.0	104	62.9	120	85.2	119	86.5
	2	5	....	5	....	8	....	8	....
		p<.33		p<.78		p<.12		p<.01	
1	0	49	65.3	46	58.7	50	79.8	51	90.5
	1	123	69.4	122	69.1	135	82.3	136	92.7
	2	78	75.8	74	73.6	86	83.7	86	97.2
	3	4	....	4	....	4	....	4	....
		p<.02		p<.00		p<.33		p<.02	
2	0	41	74.0	35	60.3	38	87.7	38	93.6
	1	76	71.9	70	70.4	79	85.4	81	93.5
	2	101	81.5	92	80.2	101	89.6	101	99.0
	3	35	83.3	31	77.8	35	93.4	35	95.5
		p<.00		p<.00		p<.20		p<.00	
3	0	24	75.5	23	68.0	26	85.1	27	89.1
	1	70	75.5	65	74.2	71	91.4	71	92.2
	2	58	77.7	53	73.3	61	92.8	61	91.8
	3	35	85.1	26	77.4	35	98.7	35	94.7
		p<.08		p<.51		p<.03		p<.13	
4&5	0	44	71.4	46	68.3	51	84.8	52	86.2
	1	110	72.1	99	74.3	117	90.9	119	84.2
	2	96	74.7	79	77.2	98	93.2	100	84.7
	3	66	80.0	47	76.3	62	96.8	63	88.5
		p<.06		p<.23		p<.00		p<.06	
6-8	0	66	68.5	69	81.1	73	88.2	75	74.3
	1	142	72.2	128	81.9	152	90.2	148	77.5
	2	96	73.5	66	88.5	94	96.2	92	80.4
	3	51	76.0	39	83.1	55	99.7	54	80.1
		p<.22		p<.22		p<.00		p<.00	
9-12	0	38	62.2	52	82.1	50	84.0	54	77.4
	1	71	65.0	70	90.9	96	89.6	95	79.4
	2	65	72.2	42	94.8	67	99.7	67	86.5
	3	22	72.7	13	80.6	26	103.7	25	90.8
		p<.05		p<.02		p<.00		p<.00	

Table 19

Mean Standard Scores For Hispanic BELEPS Students Who Entered The Program With The Lowest Level of English Proficiency (LAU-A)

Grade	Attend. Group	N	English Vocab.	N	Spanish Vocab.	N	English Reading	N	Math
K	0	72	41.7	86	61.4	101	79.7	101	77.3
	1	72	50.1	91	62.6	97	82.8	97	84.4
	2	2	....	2	....	5	....	5	....
		p<.09		p<.73		p<.08		p<.00	
1	0	28	56.9	28	60.3	27	73.7	28	86.3
	1	77	62.9	85	68.8	87	80.7	88	90.7
	2	56	72.8	59	72.4	63	81.9	63	96.7
	3	3	....	3	....	3	....	3	....
		p<.00		p<.07		p<.05		p<.01	
2	0	13	66.5	13	57.5	12	75.4	12	94.2
	1	41	67.5	45	68.3	46	80.9	48	91.8
	2	54	79.9	54	78.0	54	86.1	54	98.8
	3	18	79.1	16	74.8	18	89.6	18	94.0
		p<.01		p<.01		p<.14		p<.01	
3	0	9	75.0	9	69.9	10	79.0	12	85.4
	1	19	59.3	20	72.2	19	83.7	20	91.3
	2	14	76.0	16	70.7	17	83.7	17	90.1
	3	7	79.3	7	80.7	7	96.6	7	95.4
		p<.08		p<.54		p<.44		p<.21	
4&5	0	6	44.1	14	81.4	13	68.4	16	81.9
	1	20	65.1	29	76.6	27	83.9	29	85.0
	2	18	72.6	14	78.4	15	92.9	17	86.8
	3	10	84.6	9	76.5	11	98.6	11	93.8
		p<.01		p<.94		p<.00		p<.08	
6-8	0	15	34.9	22	84.1	21	71.7	23	71.0
	1	26	51.2	35	92.6	34	80.6	33	76.6
	2	9	62.2	9	111.9	10	96.6	10	80.7
	3	5	....	4	.....	5	....	5	....
		p<.03		p<.03		p<.00		p<.02	
9-12	0	10	33.0	25	80.9	23	68.6	27	72.1
	1	19	60.3	31	88.6	35	80.8	35	76.1
	2	10	62.1	10	111.3	9	101.1	9	81.2
	3	1	....	1	.....	2	.....	1	....
		p<.01		p<.01		p<.00		p<.01	

Table 20

Mean Standard Scores For All BELEPS Students
From Languages Other Than Spanish

Grade	Attend. Group	N	English Vocab.	N	English Reading	N	Math
K	0	5	87.0	8	94.4	8	95.6
	1	5	75.2	7	98.0	7	97.4
	2	2	....	2	....	2	....
		p<.40		p<.65		p<.84	
1	0	4	....	5	84.0	5	87.0
	1	9	69.8	11	81.5	11	90.2
	2	7	81.1	8	89.6	8	100.9
		p<.09		p<.42		p<.16	
2	0	3	....	2	....	3	.....
	1	6	68.7	6	81.3	6	91.0
	2	14	75.3	14	94.6	14	100.9
	3	3	....	3	....	3	.....
		p<.30		p<.10		p<.06	
3	0	2	....	1	.....	1	....
	1	6	75.3	6	87.2	6	89.3
	2	7	66.7	7	101.1	7	94.6
	3	5	79.4	5	100.2	5	98.0
		p<.37		p<.18		p<.37	
4&5	0	2	....	3	....	3	....
	1	13	46.0	13	77.1	13	77.6
	2	14	67.7	15	88.2	15	85.7
	3	18	75.2	17	95.8	18	90.7
		p<.00		p<.02		p<.04	
6-8	0	3	....	3	....	3	....
	1	12	50.4	12	78.8	12	77.0
	2	24	61.0	23	89.9	23	86.9
	3	11	69.6	13	94.0	13	86.7
		p<.05		p<.06		p<.06	
9-12	0	1	....	1	....	1	....
	1	14	41.2	22	83.3	21	85.3
	2	16	64.9	19	92.0	19	94.9
		p<.00		p<.12		p<.02	

Table 21

Mean Standard Scores For All Hispanic BELEPS Students by Period of Program Attendance

Grade	Attend. Group	N	English Vocab.	N	Spanish Vocab.	N	English Reading	N	Math
K	0	97	50.9	102	61.6	127	81.8	127	79.7
	1	88	56.0	103	62.3	113	84.4	112	85.9
	2			3	....	6	....		
		p<.26		p<.83		p<.13		p<.00	
1	0	44	64.1	42	60.9	44	78.8	45	90.3
	1	114	69.3	117	69.9	124	82.4	125	92.9
	2	71	75.3	71	73.1	78	83.2	78	96.8
	3	4	....	4	....	4	....	4	....
		p<.02		p<.02		p<.30		p<.05	
2	0	38	76.2	34	58.7	36	86.9	36	94.7
	1	69	72.2	69	70.5	73	85.8	75	93.7
	2	87	82.5	87	78.9	87	88.8	87	98.8
	3	32	83.6	30	77.3	32	92.2	32	94.9
		p<.00		p<.00		p<.44		p<.01	
3	0	22	77.5	22	71.0	24	84.8	26	88.7
	1	64	75.5	63	73.8	65	91.8	65	92.5
	2	51	79.2	53	73.3	54	91.7	54	91.4
	3	30	86.1	26	77.4	30	98.4	30	94.1
		p<.07		p<.77		p<.06		p<.15	
4&5	0	42	73.0	44	67.8	47	87.2	48	87.0
	1	97	75.6	99	74.3	104	92.6	106	85.0
	2	82	75.9	76	76.2	83	94.2	85	84.6
	3	48	81.8	44	76.2	45	97.2	45	87.7
		p<.16		p<.27		p<.04		p<.26	
6-8	0	63	69.1	69	81.1	70	88.5	72	73.5
	1	130	74.2	126	81.9	140	91.2	136	77.6
	2	71	77.7	65	87.8	71	98.3	69	78.3
	3	40	77.8	38	83.4	42	101.4	41	78.0
		p<.05		p<.31		p<.00		p<.01	
9-12	0	37	62.6	52	82.0	49	84.0	53	77.0
	1	57	70.8	70	90.9	74	91.5	74	77.8
	2	48	75.3	41	93.9	47	103.3	47	83.0
	3	13	76.7	13	80.6	14	107.4	13	85.5
		p<.03		p<.03		p<.00		p<.00	

Table 22

Mean Standard Scores of Children Enrolled
In Dual Language Program Models

Grade	Attend. Group	N	English Vocab.	N	Spanish Vocab.	N	English Reading	N	Math
K	0	85	47.5	95	63.5	115	81.2	115	79.1
	1	75	50.7	95	63.1	100	83.6	99	85.2
	2	4	....	4	....	5	....	5	....
		p<.50		p<.88		p<.15		p<.01	
1	0	30	53.8	31	69.1	32	75.5	32	88.1
	1	90	65.3	99	72.2	101	81.3	102	92.1
	2	53	73.0	57	74.9	62	81.4	62	96.0
	3	3	....	3	....	3	....	3	....
		p<.00		p<.36		p<.12		p<.05	
2	0	14	65.9	14	64.9	13	78.0	13	92.1
	1	42	67.6	44	68.1	47	82.5	49	93.0
	2	55	79.4	54	78.6	55	84.6	55	99.5
	3	10	79.1	9	76.3	10	90.5	10	92.4
		p<.01		p<.04		p<.50		p<.01	
3	0	7	63.9	9	68.8	8	78.8	10	78.2
	1	21	66.2	22	72.0	22	84.7	23	93.1
	2	17	77.8	18	68.3	20	89.2	20	90.7
	3	9	81.4	9	78.4	9	93.5	9	96.4
		p<.21		p<.66		p<.47		p<.00	
4&5	0	10	47.8	18	81.7	17	72.2	19	83.1
	1	21	59.6	25	81.6	29	79.3	31	83.7
	2	20	66.6	21	74.1	20	84.3	22	84.7
	3	8	77.1	8	85.8	8	88.6	9	89.1
		p<.06		p<.67		p<.11		p<.56	
6-8	0	17	44.5	25	81.8	24	74.7	26	71.5
	1	39	56.2	42	91.1	48	82.3	46	76.7
	2	14	66.5	13	95.2	15	89.1	15	74.9
	3	4	....	3	....	4	....	4	....
		p<.10		p<.32		p<.07		p<.04	
9-12	0	12	42.7	18	77.9	17	74.2	19	76.7
	1	18	67.9	29	87.8	33	83.9	32	75.6
	2	11	68.0	11	107.3	10	101.8	10	79.6
	3	0	....	0	.....	0	.....	1	....
		p<.01		p<.01		p<.00		p<.27	

Table 23

Mean Standard Scores of Children Enrolled In Modified English Instruction Program Models

Grade	Attend. Group	N	English Vocab.	N	Spanish Vocab.	N	English Reading	N	Math
K	0	16	78.6	8	39.0	18	92.3	18	91.4
	1	18	83.2	8	59.5	19	95.1	19	93.9
	2	1	....	1	....	3	....	3	....
		p<.56		p<.12		p<.60		p<.65	
1	0	19	83.4	15	37.3	18	87.3	19	94.6
	1	33	80.4	23	55.8	34	85.3	34	94.4
	2	25	81.8	17	69.5	24	89.9	24	100.2
	3	1	....	1	....	1	....	1	.....
		p<.81		p<.02		p<.51		p<.25	
2	0	27	78.1	21	57.1	25	92.8	26	94.3
	1	33	77.4	26	74.2	32	89.8	32	94.2
	2	45	84.2	37	82.6	45	95.0	45	98.4
	3	24	85.4	21	78.9	24	94.3	24	96.6
		p<.08		p<.00		p<.58		p<.21	
3	0	17	80.2	14	67.6	17	88.1	17	95.5
	1	49	79.5	43	75.4	49	94.4	48	91.8
	2	41	77.6	35	75.9	41	94.5	41	92.3
	3	26	86.5	17	76.9	26	100.5	26	94.0
		p<.15		p<.65		p<.08		p<.39	
4&5	0	34	78.3	27	59.7	33	92.5	32	88.5
	1	89	75.0	74	71.9	88	94.7	88	84.4
	2	76	76.8	58	78.3	78	95.6	78	84.8
	3	57	79.5	38	73.6	53	97.8	53	88.2
		p<.49		p<.01		p<.49		p<.07	
6-8	0	49	76.8	44	80.7	49	94.8	49	75.7
	1	103	78.2	86	77.4	104	93.9	102	77.9
	2	80	74.9	52	87.3	78	97.4	76	81.5
	3	47	77.5	36	83.2	51	99.9	50	80.1
		p<.56		p<.08		p<.19		p<.02	
9-12	0	23	73.3	25	86.0	25	92.5	25	79.6
	1	51	64.8	39	92.9	60	93.5	60	81.6
	2	54	73.1	31	90.4	57	99.4	57	87.8
	3	21	73.1	12	80.6	24	104.5	23	90.4
		p<.11		p<.32		p<.01		p<.00	